AF TD

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

OIPE IAP95 AUG 21 2006 PATENT & TRADEMARK OFFICE

Attorney Docket No. 343355600020

Group Art Unit:	2167	
Examiner:	Harold Dodds	
First Named Inventor:	R. Chu	**SUPPLEMENTAL APPEAL BRIEF**
Serial No.:	09/766,789	
Filed:	January 22, 2001	
For:	Computer-Implemented Dimension Engine	

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Mail Stop Appeal Brief- Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on
August 17, 2006 .

By Kathie J. Kopczyk

Mail Stop Appeal Brief- Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

This Supplemental Appeal Brief is submitted in response to the Notification of Non-Compliant Appeal Brief mailed July 19, 2006. Any fees due for this submission should be charged to Jones Day Deposit Account No. 501432, ref: 343355-600020.

I. Real Party in Interest

The real party in interest for this application is SAS Institute Inc., a North Carolina corporation having its principal place of business at SAS Campus Drive, Cary, North Carolina 27513. The inventors of this application have assigned their rights to SAS Institute Inc., as evidenced by documents recorded with the USPTO on January 22, 2001, at Reel 011483, Frame 0507.

II. Related Appeals and Interferences

There are no related appeals or interferences to this application.

III. Status of Claims

Claims 1-63 are pending. Of these, claims 6 and 39 are objected to. Claims 1-5, 7-38 and 40-63 are pending and being appealed.

IV. Status of Amendments

No Amendments have been filed subsequent to the present office action.

V. Summary of Claimed Subject Matter

Independent claim 1 recites a computer data store, exemplified at 32 in Figure 1 and described at page 6, lines 15-17. The computer data store stores input data that has dimension variables and at least one target variable as described at page 6, lines 15-23. A decision tree processing module, depicted at 38 in Figure 1, determines a subset of the dimension variables as described at page 7, lines 5-14 for splitting the data. The splitting

predicts the target variable as described at page 6, line 12, and page 7, line 7. The module determines the subset automatically as described at page 6, line 12. A viewer, depicted at 54 in Figure 1, generates a report using the subset and the splitting of the dimension variables, as described at page 8, lines 15-18.

Independent claim 34 recites steps in a method. One step entails storing input data that has dimension variables and at least one target variable, as described at page 6, lines 15-23. A subset of the dimension variables for splitting the input data is determined automatically wherein the splitting predicts the target variable, as described at page 6, line 12 and page 7, lines 5-14. A report using the subset and the splitting of the dimension variables is generated, as described at page 8, lines 15-18.

Independent claim 63 recites the above storing and subset determining steps. Additionally in claim 63, a request is received from a non-technical individual, as described at page 8, lines 19-21, p. 29, line 4, and page 31, lines 9-11. The subset and dimension variables are displayed so that a non-technical individual can alter which dimension variables are included in the subset, as described at page 8, lines 3-10 and 19-22. A report, for multi-dimension data analysis, is generated using the altered subset, as described at page 8, lines 11-18 and page 16, lines 22-24.

As discussed above, independent claims 1, 34 and 63 are directed to a multi-dimension data analysis techniques. Such techniques can be used to handle the large volumes of transactional data generated by enterprises that are generally stored in a data warehouse or an On-Line Analytical Processing (OLAP) system. For example, transactional data can contain information on the outcomes of enterprise operations, such as a for-profit business's records of which customers bought what products. Similarly, a

government agency may have records on which people requested what services. Likewise, a non-profit organization may have records of which donors gave money to what projects. The data sets could be so large as to have in some situations hundreds of dimension variables whose values are stored in the data store.

The claimed subject includes computer-implemented multi-dimension data analysis techniques wherein a computer data store is used to store input data that has dimension variables and at least one target variable. For example, a study may have been conducted for a company to analyze the purchasing habits of the company's customers. The study may have gathered data related to such dimension variables as the frequency that a customer has ordered from the company's catalog, the total amount the customer spent for a purchase, or what type of products the customer has purchased. An example of a target variable is a variable that indicates whether a customer has purchased an item.

A decision tree processing module accesses the data store to determine a subset of the dimension variables for splitting the input data, wherein the splitting by the dimension variable subset predicts the target variable. The decision tree processing module automatically determines the subset of the dimension variables. A multi-dimension viewer generates a report using the determined dimension variables subset and the splitting of the dimension variables.

In this manner, the claimed subject matter allows a marketing analyst (i.e., a non-technical individual) who may not be interested in the details of the decision tree algorithm to view automatically the determined data groupings with the OLAP viewer. The marketing analyst is now able to examine data that may contain hundreds of

dimensions because the data is automatically and intelligently grouped by the present invention.

VI. Grounds Of Rejection To Be Reviewed On Appeal

Claims 1-5, 7-38 and 40-63 stand rejected. Of these, independent claims 1 and 34 are rejected under 35 U.S.C. § 103(a) over Sang'udi et al. (U.S. Patent No. 6,480,194) and Anwar (U.S. Patent No. 6,750,864). Claim 63 is rejected under 35 U.S.C. § 103(a) over Anwar, Sang'udi, and Thomas (U.S. Patent No. 6,490,719).

VII. Argument

A. The Cited References (i.e., Anwar) Does Not Teach A Decision Tree Processing Module Automatically Determining A Subset Of Dimension Variables.

Claims 1-5, 7-38 and 40-63 stand rejected by the Examiner. None of the cited references, either alone or in combination, disclose a decision tree process module that automatically determines a subset of dimension variables, as required in independent claim 1.

In an Advisory Action dated November 3, 2005, the Examiner found these arguments unpersuasive. In the Advisory Action, the Examiner stated:

> ... Anwar teaches ... the decision tree processing module... "automatically determines the subset of the dimension variables" at col. 26, lines 63-65, col. 44, lines 31-34, and col. 36, lines 19-23 ....

Assignee respectfully disagrees with the Examiner's positions. Accordingly, Assignee has filed this paper with the United States Patent Office.

The Examiner's interpretation of "wherein the decision tree processing module automatically determines the subset of the dimension variables" constitutes clear error. Claim 1 is directed to a multi-dimension data analysis apparatus through use of a decision tree processing module. Claim 1 contains a computer data store that stores input data. The input data has multiple dimension variables and at least one target variable. As a non-limiting example, an input data set may contain large data sets that are associated with many dimension variables, such as those shown in Figure 2 of assignee's application (e.g., a marital dimension variable, gender dimension variable, a single mom dimension variable, etc.).

Claim 1 recites that a decision tree processing module automatically determines a subset of the dimension variables for splitting the input data. Through a decision tree processing approach, the splitting by the dimension variable subset can be used to predict the target variable.

The Advisory Action asserts that the Anwar reference teaches that a decision tree process module automatically determines the subset of dimension variables (as required by claim 1) at col. 26, lines 63-65, col. 44, lines 31-34, and col. 36, lines 19-23. These passages from Anwar are as follows:

> [(1)] Next, ACTG will evaluate all valid combinations automatically to determine the best cross-tab construct to present to the user.
> [(See Anwar at col. 26, lines 63-65.)]
>
> [(2)] In order to extract useful information (subsets of training data, statistical indices or the like) from a training set, the DMT has to perform data processing which is related to OLAP tasks.
> [(See Anwar at col. 44, lines 31-34.)]
>
> [(3)] The user can add dependent variables by grabbing a variable (dimension or member) from a list and drag-n-drop the new

> variable into the cross-tab wherever desire and the cross-tab control will add the dropped in variable to the cross-tab. The user can remove and dependent variable by simply grabbing the variable in a cross-tab and dropping outside the cross-tab.
> [(See Anwar at col. 36, lines 19-23.)]

Assignee disagrees that these excerpts from Anwar disclose automatically determining the subset of dimension variables as required by claim 1, let alone disclose that a decision tree processing module performs such an automatic determination of the subset of dimension variables. As an illustration, excerpt #1 may be discussing an automatic determination, but it is in the context of what is the best **cross-tab construct** to present to the user, and not to automatically determine through a ***decision tree*** approach a subset of dimension variables as required in claim 1. A cross-tab construct is significantly different from the decision tree subject matter of claim 1. To illustrate this, assignee notes that the Anwar reference itself mentions that "The term 'cross-tab' is a 2D view of an n-dimensional matrix." (See col. 5, lines 36-37 of Anwar). Thus the automatic generation of a cross-tab construct as defined by the Anwar reference involves significantly different subject matter from claim 1's subject matter which involves generation of a subset of dimension variables through a decision tree approach.

Excerpt #2 of Anwar does not disclose any automatic determination, let alone an automatic determination of the subset of dimension variables of claim 1. Rather excerpt #2 of Anwar is only disclosing that training sets are difficult for OLAP databases and how to extract useful information from a training set.

Excerpt #3 of Anwar also does not disclose any automatic determination, let alone an automatic determination of the subset of dimension variables of claim 1. In fact this excerpt further evidences the manual approach of Anwar by disclosing

> [(3)] The **user** can add dependent variables by **grabbing** a variable (dimension or member) from a list and **drag-n-drop** the new variable into the cross-tab wherever desire and the cross-tab control will add the dropped in variable to the cross-tab. The **user** can remove and dependent variable by **simply grabbing** the variable in a cross-tab and **dropping** outside the cross-tab. (At col. 36, lines 19-23; Emphasis added).

In excerpt #3, the **user** is performing manual actions, such as **grabbing**, **dropping**, and **drag-n-drop** actions with respect to variables.

The Anwar reference does not disclose a decision tree that determines a subset of the dimension variables for splitting the input data as required by claim 1 in combination with its other limitations. Instead, the Anwar reference discloses a user, though a manual process, selecting variables: "First, ***the user selects*** one or more dependent variables and a plurality of independent variable from a list of dimensions and members associated with a multidimensional database (MDD)." (See col. 33, lines 26-29; emphasis added.)

Moreover, the Anwar reference is concerned with a different problem than what the subject matter of claim 1 is addressing. For example, the Anwar reference is concerned about using "a decision tree generator where the number of dependent variables is greater than one." (See col. 32, lines 65-67.) The Anwar reference goes into more detail about this as follows:

> In a traditional decision tree, the top of the tree is a single dependent variable or decision and the resulting decision tree shows all independent variables and their values that relate to the dependent variable. However, traditional decision trees are not designed to handle more than one dependent variable. On the other hand, the decision tree generator of the present invention is specifically designed to handle two or more dependent variables and provide for efficient visualization of the multi-dependent decision trees using novel graphic constructs.
> (See col. 32, line 65 - col. 33, line 8)

Accordingly, the Anwar reference is concerned with a different problem than what claim 1 is directed.

As shown in the analysis of the cited excerpts of Anwar, Anwar does not disclose the limitations of claim 1, such as a decision tree process module that automatically determines the subset of dimension variables as required by claim 1 in combination with its other limitations. Because of such differences, Anwar (whether considered alone or in combination with the other cited references) does not render claim 1 obvious and thus claim 1 is allowable and should proceed to issuance.

Claim 34 is directed to a computer-implemented multi-dimension data analysis method. Claim 34 recites in combination with its other limitations that a subset of the dimension variables is automatically determined. Because the cited references (whether viewed alone or in combination) do not teach, disclose or suggest such limitations of claim 34, claim 34 and its dependent claims are allowable.

Claim 63 is directed to a computer-implemented method for multi-dimension data analysis by a non-technical individual. Claim 63 recites in combination with its other limitations that a subset of the dimension variables is automatically determined. Because the cited references (whether viewed alone or in combination) do not teach, disclose or suggest such limitations, claim 63 is allowable.

For the above reasons, Applicant respectfully submits that the pending claims are allowable, and requests the withdrawal of the rejections.

VIII. Claims Appendix

An appendix is attached hereto setting forth a copy of the pending claims involved in the appeal.

IX. Evidence Appendix

No evidence has been entered and relied upon.

X. Related Proceedings Appendix

There are no related proceedings related to this application.

Respectfully submitted,

By: John Biernacki

John V. Biernacki
Reg. No. 40,511
JONES DAY
North Point
901 Lakeside Avenue
Cleveland, Ohio 44114
(216) 586-3939

CLAIMS APPENDIX

1. **(PREVIOUSLY PRESENTED)** A computer-implemented multi-dimension data analysis apparatus, comprising:

a computer data store for storing input data that has dimension variables and at least one target variable;

a decision tree processing module connected to the data store that determines a subset of the dimension variables for splitting the input data, wherein the splitting by the dimension variable subset predicts the target variable; and

wherein the decision tree processing module automatically determines the subset of the dimension variables;

a multi-dimension viewer that generates a report using the determined dimension variables subset and the splitting of the dimension variables.

2. **(ORIGINAL)** The apparatus of claim 1 wherein the dimension variables subset includes continuous variables.

3. **(ORIGINAL)** The apparatus of claim 1 wherein the dimension variables subset includes category-based variables.

4. **(ORIGINAL)** The apparatus of claim 1 further comprising:

a selector module so that a user can alter which dimension variables to include in the subset.

5. **(ORIGINAL)** The apparatus of claim 4 wherein at least one statistic measure is provided to the user that is indicative of how well the splitting of the dimension variables predicts the target variable.

6. **(ORIGINAL)** The apparatus of claim 5 wherein the statistic measure is a logworth statistic measure.

7. **(ORIGINAL)** The apparatus of claim 1 further comprising:

a selector module so that a user can alter values at which the input data is split by the decision tree processing module.

8. **(ORIGINAL)** The apparatus of claim 1 wherein the input data set includes a plurality of dimension variables and a single target variable.

9. **(ORIGINAL)** The apparatus of claim 1 wherein the input data set includes a plurality of dimension variables and a plurality of target variables.

10. **(ORIGINAL)** The apparatus of claim 1 wherein the decision tree processing module splits the input data into groups, wherein the multi-dimension viewer generates a report using the groups.

11. **(ORIGINAL)** The apparatus of claim 10 wherein the decision tree processing module uses a competing initial splits approach to determine a subset of the dimension variables.

12. **(ORIGINAL)** The apparatus of claim 11 wherein an initial split variable is indicated as most important variable in predicting the target variable.

13. **(ORIGINAL)** The apparatus of claim 12 wherein a second split variable is indicated as second most important variable in predicting the target variable.

14. **(ORIGINAL)** The apparatus of claim 1 wherein the decision tree processing module generates binary splits of the input data.

15. **(ORIGINAL)** The apparatus of claim 1 wherein the decision tree processing module generates splits of the input data that are other than binary splits.

16. **(ORIGINAL)** The apparatus of claim 1 wherein the generated report is viewed substantially adjacent to the dimension variables subset and the splitting values of the dimension variables subset.

17. **(ORIGINAL)** The apparatus of claim 1 wherein the report has a format selected from the group consisting of a textual report format, tabular report format, graphical report format, and combinations thereof.

18. **(ORIGINAL)** The apparatus of claim 17 wherein a marketing analyst selects one of the report formats in order to view the determined dimension variables subset and the splitting of the dimension variables.

19. **(ORIGINAL)** The apparatus of claim 18 wherein the input data includes more than fifty dimension variables, wherein the determined dimension variables subset includes less than seven dimension variables that are viewed by the marketing analyst.

20. **(ORIGINAL)** The apparatus of claim 1 wherein a user selects a type of summary statistics to view the determined dimension variables subset and the splitting of the dimension variables.

21. **(ORIGINAL)** The apparatus of claim 1 further comprising:

a model repository for storing a model that contains the dimension variables and splitting values of the dimension variables.

22. **(ORIGINAL)** The apparatus of claim 21 wherein the decision tree processing module splits the input data into a first set of groups according to first splitting rules to form a first model,

wherein the decision tree processing module splits different input data into a second set of groups according to second splitting rules to form a second model,

wherein the model repository includes a splitting rules index to store which splitting rules are used with which model.

23. **(ORIGINAL)** The apparatus of claim 22 wherein the splitting rules index is searched in order to locate a model stored in the model repository.

24. **(ORIGINAL)** The apparatus of claim 23 wherein the model repository includes a project level storage means, a diagram level storage means, and a model level storage means for storing the first and second models.

25. **(ORIGINAL)** The apparatus of claim 22 wherein a search request is provided over a computer network to retrieve the first model from the model repository.

26. **(ORIGINAL)** The apparatus of claim 25 wherein the computer network is an Internet network.

27. **(ORIGINAL)** The apparatus of claim 22 wherein the model repository includes a plurality of specialty splitting rules indices that are used to locate a model stored in the model repository.

28. **(ORIGINAL)** The apparatus of claim 27 wherein the specialty splitting rules indices are indices selected from the group consisting of marketing specialty splitting rules indices, sales specialty splitting rules indices, and combinations thereof.

29. **(ORIGINAL)** The apparatus of claim 22 wherein the model repository includes a mini-index means with a connection to the splitting rules index.

30. **(ORIGINAL)** The apparatus of claim 1 wherein a data mining application provides construction of a process flow diagram, wherein the process flow diagram includes nodes representative of the input data and a variable configuration module.

31. **(ORIGINAL)** The apparatus of claim 30 wherein an activated variable configuration module node provides a graphical user interface within which a user can alter which dimension variables to include in the subset.

32. **(ORIGINAL)** The apparatus of claim 31 wherein the process flow diagram further includes a node representative of the decision tree processing module that has a competing initial splits approach for determining the subset of the dimension variables.

33. **(ORIGINAL)** The apparatus of claim 31 wherein the process flow diagram further includes a node representative of the decision tree processing module that has a non-competing initial splits approach for determining the subset of the dimension variables.

34. **(PREVIOUSLY PRESENTED)** A computer-implemented multi-dimension data analysis method, comprising the steps of:

storing input data that has dimension variables and at least one target variable;

determining a subset of the dimension variables for splitting the input data, wherein the splitting using the dimension variable subset predicts the target variable; and

wherein the subset of the dimension variables is automatically determined;

generating a report using the determined dimension variables subset and the splitting of the dimension variables.

35. **(ORIGINAL)** The method of claim 34 wherein the dimension variables subset includes continuous variables.

36. **(ORIGINAL)** The method of claim 34 wherein the dimension variables subset includes category-based variables.

37. **(ORIGINAL)** The method of claim 34 further comprising the step of:

altering which dimension variables to include in the subset of the dimension variables.

38. **(ORIGINAL)** The method of claim 37 further comprising the step of:

providing at least one statistic measure that is indicative of how well the splitting of the dimension variables predicts the target variable.

39. **(ORIGINAL)** The method of claim 38 wherein the statistic measure is a logworth statistic measure.

40. **(ORIGINAL)** The method of claim 34 further comprising the step of:

altering values at which the input data is split.

41. **(ORIGINAL)** The method of claim 34 wherein the input data set includes a plurality of dimension variables and a single target variable.

42. **(ORIGINAL)** The method of claim 34 wherein the input data set includes a plurality of dimension variables and a plurality of target variables.

43. **(ORIGINAL)** The method of claim 34 further comprising the step of:

using a decision tree algorithm to determine the subset of the dimension variables by which to split the input data.

44. **(ORIGINAL)** The method of claim 43 wherein the decision tree algorithm splits the input data into groups, wherein the multi-dimension viewer generates a report using the groups.

45. **(ORIGINAL)** The method of claim 44 wherein the decision tree algorithm uses a competing initial splits approach to determine the subset of the dimension variables.

46. **(ORIGINAL)** The method of claim 45 wherein an initial split variable is indicated as most important variable in predicting the target variable.

47. **(ORIGINAL)** The method of claim 46 wherein a second split variable is indicated as second most important variable in predicting the target variable.

48. **(ORIGINAL)** The method of claim 34 further comprising the step of:

generating binary splits of the input data.

49. **(ORIGINAL)** The method of claim 34 further comprising the step of:

generating splits of the input data that are other than binary splits.

50. **(ORIGINAL)** The method of claim 34 wherein the generated report is viewed substantially proximate to the dimension variables subset and the splitting values of the dimension variables subset.

51. **(ORIGINAL)** The method of claim 34 wherein the report has a format selected from the group consisting of a textual report format, tabular report format, graphical report format, and combinations thereof.

52. **(ORIGINAL)** The method of claim 51 wherein a marketing analyst selects one of the report formats in order to view the determined dimension variables subset and the splitting of the dimension variables.

53. **(ORIGINAL)** The method of claim 52 wherein the input data includes more than fifty dimension variables, wherein the determined dimension variables subset includes less than seven dimension variables that are viewed by the marketing analyst.

54. **(ORIGINAL)** The method of claim 34 wherein a user selects a type of summary statistics to view the determined dimension variables subset and the splitting of the dimension variables.

55. **(ORIGINAL)** The method of claim 34 further comprising the step of:

storing a model in a model repository, wherein the model contains the dimension variables and splitting values of the dimension variables.

56. **(ORIGINAL)** The method of claim 55 further comprising the step of:

storing the model in a project level storage means, a diagram level storage means, and a model level storage means of the model repository.

57. **(ORIGINAL)** The method of claim 55 wherein a search request is provided over a computer network to retrieve the model from the model repository.

58. **(ORIGINAL)** The method of claim 57 wherein the computer network is an Internet network.

59. **(ORIGINAL)** The method of claim 55 wherein the model repository includes a plurality of specialty splitting rules indices that are used to locate the model stored in the model repository.

60. **(ORIGINAL)** The method of claim 59 wherein the specialty splitting rules indices are indices selected from the group consisting of marketing specialty splitting rules indices, sales specialty splitting rules indices, and combinations thereof.

61. **(ORIGINAL)** The method of claim 34 wherein a data mining application provides construction of a process flow diagram, wherein the process flow diagram includes nodes representative of the input data and a variable configuration module.

62. **(ORIGINAL)** The method of claim 61 further comprising the step of:

activating the variable configuration module node so that a user can alter which dimension variables to include in the subset.

63. **(PREVIOUSLY PRESENTED)** A computer-implemented method for multi-dimension data analysis by a non-technical individual, comprising the steps of:

storing input data that has dimension and target variables;

receiving a request from the non-technical individual to analyze the stored input data;

after receiving the request, determining a subset of the dimension variables for splitting the input data, wherein the splitting using the dimension variable subset predicts the target variable;

wherein the subset of the dimension variables is automatically determined;

displaying the determined dimension variables subset and the dimension variables so that the non-technical individual can alter which of the dimension variables are included in the dimension variables subset; and

generating a report for the non-technical personnel using the dimension variables subset as altered by the non-technical individual,

whereby the generated report is used for multi-dimension data analysis by the non-technical individual.

<u>Evidence Appendix</u>

No evidence has been entered and relied upon.

Related Proceedings Appendix

There are no related proceedings related to this application.